Working Paper No. 93/2

X11 TIME SERIES DECOMPOSITION

AND SAMPLING ERRORS

Andrew Sutcliffe

December 1993

ABSTRACT

There is considerable interest in analysing the effect of sample design on time series decomposition into components such as trend, seasonal and residual. One of the main deficiencies in past analysis has been a lack of a realistic model of an officially used time series decomposition method. This paper demonstrates a realistic approximation for X11 and shows how such a model can be used to analyse the effect of sample design on time series decomposition.

TIME SERIES DECOMPOSITIONUSING X11 AND SAMPLING ERRORS

1. Introduction.

There is considerable interest in the effects of sampling error on the components of time series decomposition, such as, the seasonally adjusted, trend and residual data. For example it is useful to have standard errors on the seasonally adjusted data, trend, seasonal and residual which represent the effects of the sampling process. An important question is what proportion of the observed variability in a time series is due to sampling error. Then there are design considerations, for example given a certain sample design and characteristics what is an "optimal" time series decomposition. One of the main deficiencies in past analysis has been providing a realistic model of an officially used time series decomposition method. This paper demonstrates a realistic model for X11(Shiskin et al, 1967) and shows how such a model can be used to analyse the effect of sampling design on time series decomposition.

To give a appreciation of the complexity of the problem consider the simple time series model

1.1 $$\tilde{y} = \tilde{t} + \tilde{s} + \tilde{e}$$

where

1.2 $\tilde{t}$ is the trend
$\tilde{s}$ is the seasonal
$\tilde{e}$ is the residual/irregular

It is reasonable to assume the residual contains the following components

1.3 $$\tilde{e} = er\tilde{w} + e\tilde{s} + en\tilde{s}$$

where

1.4 $er\tilde{w}$ is the "real world" noise
$e\tilde{s}$ is the sampling error
$en\tilde{s}$ is the non-sampling error

However the following sources of error may apply to the "true" decomposition :

(i) The sampling error will affect the estimation of the trend, seasonal and "real world" noise. The sampling error may also contain "trend like" and seasonal behaviour.

(ii) The non-sampling error will affect the estimation of the trend, seasonal and "real world" noise. In addition the non-sampling error may well contain trend and seasonal components itself.

(iii) Revision to the data will effect components to varying degrees.

(iv) The components at the end of the data will be revised as more data becomes available due to the X11 method. That is assuming the central filters in X11 give the "true" components the estimates at the end of the data will only be an approximation.

(v) It is assumed that X11 methodology can adequately estimate the trend, seasonal and residual.

(vi) It is assumed that other components such as trading-day, moving holidays, and abrupt changes in the level or seasonal pattern are adequately estimated when they exist in the data.

2. Model to analyse the effect of sampling error on time series decomposition

The Australian Bureau of Statistics (ABS) has used the X11 method for all officially seasonally adjusted time series since 1967. Generally time series sample surveys have error estimates that are correlated. Hence the two main tools to enable an analysis of the effect of sampling errors on a time series are a realistic model of X11 and a covariance matrix associated with the sample design.

3. A realistic model of X11.

3.1 Previous models

Young (1963) gave a basic linear approximation to X11. In general though this, and other subsequent work has been inadequate to accurately represent X11 as used in practice. The main area's where the previous models have been inadequate are :-

(i) Studies have mainly concentrated on the central filters used in X11. These are not representative of the actual filters used by X11 (for example compare graph A3 with A5).

(ii) In the main the standard options available in X11 have been used. All other options have been ignored including the fact that X11 uses different filters for short time series (e.g. 5 years long).

(iii) It is not recognised that logging the data and adjusting additively is not equivalent to the multiplicative option in X11 especially in terms of the arithmetic levels.

(iv) The modification for outliers used in X11 is usually completely ignored.

(v) Effects such as trading-day, moving holidays, trend breaks, seasonal breaks etc. have been ignored.

For example the often quoted paper by Cleveland et al. (1976), is clearly deficient on all of the above concerns. Other papers that are deficient in at least some of the considerations above include Hausman and Watson (1985), Maravall (1985), and Butter and Mourik (1990). A much more comprehensive treatment is given in the recently published working paper by Dagum, Chhab and Chiu (1993).

3.2 Other Methods

Some authors have assumed other theoretical models for the decomposition. These include SABL,STL (Cleveland et al.) which use moving medians and local regression respectively, BSM (Maravall et al. 1985) a basic structural model , and STM(Butter and Mourik, 1990) a structural model using a State Space representation and estimated using the Kalman filter. This approach ignores the fact that many major statistical organisations are using X11. In addition these alternative models are still deficient with respect to some of the points listed above.

Professor James Durbin said in a speech in 1986 "...and a great deal of progress has been made in the theory of time series analysis. It would, therefore, have been natural to expect that improvements in methods of seasonal adjustment would have taken place at a rapid rate, and that the techniques in daily use today would have been revolutionised compared to the X11 method of twenty year ago. However this has not happened." Part of the problem even today is that while many new structural models have been proposed and offer promising alternatives to X11 they are not developed or tested to a form suitable for a large statistical organisation to decompose hundreds, or even thousands of time series with differing characteristics.

3.3 *Non-iterative version of X11.*

One of the design features of X11 is to provide an ad-hoc iterative procedure that converges to a relatively stable estimate quickly. A non-iterative procedure that gives very similar results to X11 is outlined below. A feature of this implementation is that a general matrix language (PROC IML in SAS) has been used. This has enabled compact code and all of the options available in X11 (and some that are not) to be included.

3.4. Notation

3.4.1 n by 1 vectors where n is the length of the data

$\tilde{y}$ original data
$\tilde{t}$ trend
$\tilde{s}$ seasonal
$\tilde{e}$ residual/irregular
$\tilde{w}$ weights(0 no modification, 1 fully modified)
$\tilde{a}$ seasonally adjusted data

3.4.2 n by n matrices :

T A matrix to derive the trend from seasonally adjusted data. For example for a 7 term Henderson the matrix would look like:

$$T = \begin{bmatrix} 0.535 & 0.383 & 0.116 & -0.034 & 0 & . & . & . & 0 \\ 0.289 & 0.410 & 0.294 & 0.061 & -0.054 & 0 & . & . & 0 \\ 0.034 & 0.275 & 0.399 & 0.287 & 0.058 & -0.053 & 0 & . & 0 \\ -0.059 & 0.059 & 0.294 & 0.412 & 0.294 & 0.059 & -0.059 & . & 0 \\ 0 & . & . & . & . & . & . & . & . \\ 0 & . & . & . & . & . & . & . & . \\ 0 & . & . & . & . & . & . & . & . \\ 0 & . & . & . & . & . & . & . & . \\ 0 & . & . & . & . & -0.034 & 0.116 & 0.383 & 0.535 \end{bmatrix}$$

3.4.3 $S1$ A matrix to compute the seasonal factors from the detrended data. For example a 3 by 5 moving average.

3.4.4 $S2$ A matrix to correct for levels.. For example a 2 by 12 moving average.
Let

3.4.5 $S = (I - S2) * S1$

Then for a simple additive model

3.4.6 $\tilde{y} = \tilde{t} + \tilde{s} + \tilde{e}$

we have three equations

3.4.7
$$\begin{aligned} \tilde{t} &= T * (\tilde{y} - \tilde{s}) \\ \tilde{s} &= S * (\tilde{y} - \tilde{t}\,) \\ \tilde{e} &= \tilde{y} - \tilde{t} - \tilde{s} \end{aligned}$$

Hence

3.4.8
$$\begin{aligned} \tilde{t} &= T * (\tilde{y} - S * (\tilde{y} - \tilde{t}\,)) \\ (I - T * S) * \tilde{t} &= (T - T * S) * \tilde{y} \end{aligned}$$

Providing that the inverse exists we have

3.4.9 $\tilde{t} = (I - T * S)^{-1} * (T - T * S) * \tilde{y}$

Hence the component's $\tilde{t}, \tilde{s}, \tilde{e}$ and the adjusted data $\tilde{a}$ can be directly computed from $\tilde{y}$.

3.4.10
$$\begin{aligned} \tilde{t} &= \hat{T} * \tilde{y} && \text{where} && \hat{T} = (I - T * S)^{-1} * (T - T * S) \\ \tilde{s} &= \hat{S} * \tilde{y} && \text{where} && \hat{S} = S - S * \hat{T} \\ \tilde{e} &= \hat{E} * \tilde{y} && \text{where} && \hat{E} = I - \hat{T} - \hat{S} \\ \tilde{a} &= \hat{A} * \tilde{y} && \text{where} && \hat{A} = (I - \hat{S}) \end{aligned}$$

3.5. Multiplicative adjustment
If

3.5.1 $\tilde{y} = \tilde{t} * \tilde{s} * \tilde{e}$

then logging 3.3.1 gives an additive model in logs

3.5.2 $\log(\tilde{y}) = \log(\tilde{t}) + \log(\tilde{s}) + \log(\tilde{e})$

Hence estimates of the components are given by

3.5.3
$$
\begin{aligned}
\tilde{t} &= \exp(\hat{T} * \log(\tilde{y})) \\
\tilde{s} &= \exp(\hat{S} * \log(\tilde{y})) \\
\tilde{e} &= \exp(\hat{E} * \log(\tilde{y}))
\end{aligned}
$$

Unfortunately this leads to geometric moving averages being applied. This means that levels are maintained in geometric rather than arithmetic terms (as say multiplicative X11 does). It is interesting to note that the symmetric Henderson weights do have the property that the weights applied to logged data and then un-logged are very close to the weights applied directly to the un-logged data (because Henderson moving averages leave polynomial trends of up to cubics unchanged). The end weights and the moving averages used to obtain the seasonal factors from the de-trended data do not have this property. Hence, if a seasonal series is adjusted using multiplicative X11 and compared to the same adjustment logging the data, applying an additive adjustment and then un-logging the data a systematic bias is observed in say the trend levels. There are many ways to attempt to correct for this bias. A simple and effective way is to correct the seasonal factors obtained with the log additive model for multiplicative level bias.

That is

3.5.4
$$
\begin{aligned}
sbia\tilde{s} &= \exp(\hat{S} * \log(\tilde{y})) \\
\tilde{s} &= sbia\tilde{s} / (S2 * sbia\tilde{s})
\end{aligned}
$$

3.6. Modifications for extremes

An important part of X11 is to deal with outliers. To accurately represent X11 (and to provide good estimates of the components) outliers must be modified. A simple and effective method is to use the concept of a weight applied to the data.

That is, let $\tilde{w}$ be a vector of weights to apply, where $w(i) = 0$ gives no modification and $w(i) = 1$ gives full modification, and $0 < w(i) < 1$ gives partial modification. Then it can be shown that the matrices to derive the trend, seasonal and residual can be modified to allow for outliers as shown below.

Assuming $\hat{T}, \hat{S}$ and $\hat{E}$ have been computed as outlined above then they can be modified to take account of outliers as follows

3.6.1
$$
\begin{aligned}
\hat{E} &= (I - (I - \hat{E}) * diag(\tilde{w}))^{-1} * \hat{E} \\
\hat{T} &= \hat{T} * (I - diag(\tilde{w}) * \hat{E}) \\
\hat{S} &= \hat{S} * (I - diag(\tilde{w}) * \hat{E})
\end{aligned}
$$

The weights $\tilde{w}$ can be taken from X11 or directly computed using a method similar to X11. In theory the computation of the weights would need to be taken into account for standard errors on the estimates. This has not been attempted in this paper.

Several alternative methods of handling outliers are currently being investigated. These include modifications of the form

3.6.2 $$\tilde{y} - \tilde{\lambda}$$

where $\tilde{\lambda}$ is of the form

3.6.3 $$\tilde{\lambda} = g * (I + g * \hat{E}' * \hat{E})^{-1} * \hat{E}' * \hat{E} * \tilde{y} \text{ where } 0 \le g < \infty$$

or with suitable G

3.6.4 $$\tilde{\lambda} = g * (I + g * G' * \hat{E}' * \hat{E} * G)^{-1} * G' * \hat{E}' * \hat{E} * \tilde{y}$$

One interesting application is the choice of G to enable outliers to receive more smoothing rather that being removed or reduced in the data.

3.7. Trading-day, moving holidays and other influences

Others types of components such as trading-day, moving holidays and abrupt changes in the level or seasonal pattern can be incorporated into the methodology outlined above.

For example letting X be an n by k matrix where k is the number of parameters to be estimated and X appropriately formulated for trading-day (Young 1965) or moving holidays. Then these components can be incorporated as

3.7.1 $$\hat{D} = (I - H * (I - \hat{E}))^{-1} * H * \hat{E}$$

where $H = X * (X' * X)^{-1} * X'$.

The other components are modified as

3.7.2 $$\hat{T} = \hat{T} * (I - \hat{D})$$
$$\hat{S} = \hat{S} * (I - \hat{D})$$
$$\hat{E} = I - \hat{T} - \hat{S} - \hat{D}$$

3.8. Extensions to X11

The basic X11 algorithm can easily be extended to include other models, for example it is a simple extension to allow for regression estimates of the trend and seasonal components. The residual component could be allowed to follow a moving average process (Box-Jenkins). For example for the basic model

3.8.1 $$\tilde{y} = \tilde{t} + \tilde{s} + \tilde{e}$$

the residual $\tilde{e}$ could follow a moving average process of order 1

3.8.2 $e(t) = \varepsilon(t) - \theta * \varepsilon(t-1)$ where $\varepsilon \sim N(0, \sigma^2)$

letting

3.8.3
$$G = \begin{bmatrix} 1 & 0 & 0 & 0 & . & 0 \\ -\theta & 1 & 0 & 0 & . & 0 \\ 0 & -\theta & 1 & 0 & . & 0 \\ 0 & . & . & . & . & . \\ . & . & . & -\theta & 1 & 0 \\ 0 & . & 0 & 0 & -\theta & 1 \end{bmatrix}$$

we have

3.8.4
$$\tilde{\varepsilon} = (I - \hat{E} * (I - G))^{-1} * \hat{E} * \tilde{y}$$

Hence θ could be estimated by least squares by minimising $\tilde{\varepsilon}' * \tilde{\varepsilon}$ or with appropriate modifications by maximum likelihood.

3.9. Implementation

The vector and matrix formulation given above is ideally suited to Proc IML in SAS, and other matrix oriented packages. All computations have been done using IML. The computations for the Henderson moving averages have been done using algorithms for the central and surrogate weights (used at the ends of the data) to make the method more general.

As an example Australian Monthly Retail Turnover has been used. The data from the 1992 annual re-analysis with the same prior correction factors has been used. The span of the analysis has been restricted to 7/83 to 6/92 due to the large amount of computer memory required.

The seasonal factors are computed using the above methodology from 7/83 to 6/92 and the forward factors computed using the same method as in X11. The adjusted series is compared to the published figures (April 1993) in graph A1. The movements are compared in Graph A2. For this example the differences are minor. Several other time series have been looked at with similar results. More comprehensive testing is envisaged in the future.

Graphs A3 to A14 shows the weights for some selected point in time for the trend, seasonal and irregular components. Some of the features are that the weighting patterns at the ends of the data are completely different to those at the centre (e.g. compare A3 with A5). The modified weighting patterns may be similar to the unmodified in some cases, and different in others, depending on the weights and the location of the extremes (e.g. compare A7 and A8).

Graph A9 shows $\hat{T}$ graphically for the case of a 9 term Henderson, 3x5 seasonal and modifications for extremes for the Monthly Retail Turnover data. The left hand axis is the

actual weights, the bottom axis is the time point at which the filter given by the right axis is applied.

While not shown in this paper the spectral properties of the filters outlined above can be extensively analysed using the gain and phase outputs of a linear filter.

4. A covariance matrix associated with the sample design

The computation of the covariance matrix associated with the sampling design is mainly a sampling problem. There seems to be three approaches possible. These are :

(i) Compute the whole covariance matrix directly using the sampling design. There would usually be computational difficulties in computing such a matrix going back several years.

(ii) Find a model to estimate the covariances. This was the approach taken by Steel and De Mel (1987) for the Australian Labour Force Data where a "geometric decay" model is used.

(iii) Make up the covariance matrix using an educated guess for the model.

It should be noted that any covariance matrix used must be positive semi-definite to ensure zero or positive variances.

The analysis would be considerably complicated if the sampling error, non-sampling error and "real world" error are correlated. There is also the somewhat philosophical problem of looking at the "true" observed values as fixed, versus their own distribution, that is are the sample totals fixed or random variables. Given these tools the type of analysis that can be achieved is outlined below.

The basic theoretical framework is outlined in section 3. Ignoring a correction for levels, the components of a multiplicative time series analysis, with modification for extremes (using the modifications in 3.4), is given by

4.1
$$\begin{aligned} \tilde{t} &= \exp(\hat{T} * \log(\tilde{y})) \\ \tilde{s} &= \exp(\hat{S} * \log(\tilde{y})) \\ \tilde{e} &= \exp(\hat{E} * \log(\tilde{y})) \end{aligned}$$

5. Additive verses multiplicative standard errors.

The published standard errors are usually given as an additive standard error. Given the multiplicative option in X11 is almost always used this means that the basic model is given by

5.1
$$\tilde{y} = \tilde{t} * \tilde{s} * er\tilde{w} * en\tilde{s} + e\tilde{s}$$

This presents some problems, namely:

(i) Logging the data does not give a model which is easy to analyse.

(ii) Such a model cannot be totally reasonable since it implies negative values are possible in data that cannot be usually negative.

(iii) The standard errors are usually relative to the level of the data. For example the published levels, standard errors and percentage standard errors for Australian Monthly Retail Turnover is given in the table below

5.2 TABLE 1

Date	Level	Standard Error	% Standard Error
02/92	7,106.2	78.9	1.11
03/92	7,475.9	73.6	0.98
04/92	7,694.5	73.5	0.96
05/92	7,797.3	74.8	0.96
06/92	7,547.9	63.4	0.84
07/92	7,819.7	62.4	0.8
08/92	7,461	59.4	0.8
09/92	7,745.6	60.4	0.78
10/92	8,252.3	64.4	0.78
11/92	8,126.9	64.8	0.8
12/92	10,637.7	91.7	0.86
01/93	7,789.5	67.2	0.86
02/93	7,108	59.5	0.84
03/93	7,831.6	66.2	0.85
04/93	7,905.4	66.5	0.84

Because of the above problems it is more reasonable to have the covariance matrix in terms of a multiplicative error. That is

5.3 $$\tilde{y} = \tilde{t} * \tilde{s} * er\tilde{w} * en\tilde{s} * (1 + e\tilde{s})$$

6. Variance on the estimated components

Assuming a relative covariance matrix C, then variances of the components are given by

6.1 $$\sigma^2_{es}(\tilde{t}) = diag(\hat{T} * C * \hat{T}')$$
$$\sigma^2_{es}(\tilde{s}) = diag(\hat{S} * C * \hat{S}')$$

$$\sigma_{es}^2(\tilde{e}) = diag(\hat{E} * C * \hat{E}^{'})$$

Hence 95 per cent confidence bounds (assuming ±2 sigma is 95%) for the components are given by

6.2
$$\begin{array}{l} \tilde{t} * (1 \pm 2 * \sigma_{es}(\tilde{t})) \\ \tilde{s} * (1 \pm 2 * \sigma_{es}(\tilde{s})) \\ \tilde{e} * (1 \pm 2 * \sigma_{es}(\tilde{e})) \end{array}$$

From these formulas additive errors can be numerically computed. It should be noted that these estimates are highly correlated, and hence cannot be used to give simultaneous confidence bounds on several time points. In addition these formulas will only provide estimates of the uncertainly due to the sampling error.

To give an example Monthly Retail Turnover data has been analysed. It should be noted that this analysis is for demonstration purposes only, and is not necessarily realistic. In practice the covariance matrix associated with the sample design would include explicit allowance for the rotation used, sample redesign and other known sampling design characteristics. The sample design area of the ABS is currently researching such covariance matrices for the ABS sample surveys.

An examination of table 1 above and the standard errors on the movements given in the publication a standard error of about 1 per cent and a high correlation between successive time points might be a reasonable model for the covariance matrix. A possible model might be an AR(1) model of the form

6.3
$$es_t = \rho * es_{t-1} + \varepsilon_t \text{ where } \varepsilon_t \text{ iid with variance } \sigma^2$$

it can be shown that this has the covariance matrix

6.4
$$\frac{\sigma^2}{1-\rho^2}\begin{bmatrix} 1 & \rho & \rho^2 & . & . & \rho^{n-1} \\ \rho & 1 & \rho & . & . & \rho^{n-2} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ \rho^{n-2} & . & . & \rho & 1 & \rho \\ \rho^{n-1} & . & . & \rho^2 & \rho & 1 \end{bmatrix}$$

Hence letting $\rho = 0.8$ and $\sigma^2 = 0.0001 * (1 - 0.8^2)$ gives a 1 per cent standard error and correlation at lag 1 of 0.8 (an alternative model might have been a MA(1)).

Graphs 1,2 ,3 and 4 show the standard error of the trend, seasonal factors, irregular and seasonally adjusted components as a proportion of the original standard error (1%). For example Graph 1 shows that the trend from X11 is about 104 per cent of the standard error on the original data at the end of the data, and about 90 per cent in the middle. These results would change if the options in X11 are varied, the covariance matrix is different (either different parameter values, or different model). It should be noted that the trend standard error is for the X11 trend, which in the case of Retail is different from the

published trend. This is because ABS currently uses a 9 term Henderson for the seasonal adjustment, while the published trend is uses a 13 term Henderson, and no allowance is made for outliers in the published trend.

It is relatively straightforward to modify the procedure to allow for these differences. The approximate standard error of the published trend as a proportion of the original standard error is shown in graph 5. If the proportions of the trend in graph 1 and graph 5 are compared it will be noticed that it is lower in graph 5 (NB 13 vs 9 Henderson). However this does not however imply that estimation of the trend without modifications for extremes is superior. There are two competing factors in estimating the trend (or any other component) "bias" and "variance". That is the trend produced without modifications for extremes may have a lower variance but be also very biased in producing what is deemed a reasonable trend.

A similar situation is faced at the end of the data. It has been noticed for some time series that the trend produced at the end of the data seems to be biased when compared to the final trend. Such bias could be eliminated by applying the same criteria that are used to compute the central Henderson weights at the end of the data. The result would be a much higher variance on the trend at the end of the data.

Comparing these results with the work of Steel and De Mel (op. at.) shows that while there are broad similarities in the magnitude, it is clear that there is considerably more complexity in the filters actually used in X11 to obtain the components. Some interesting features are the pronounced rises in the early years of the proportion for the seasonal factors. This is due to a predominance of full modification for outliers in those months (compare with graph 6).

The variability of the residual due to the sampling process can be compared with the estimate of variance for the observed residual. This includes "real world" error, sampling error and non-sampling error, and (assuming a constant variance), is given by :

6.5 $$\sigma_e^2 = \frac{1}{n} \Sigma_{t=1}^{n} (e_t - 1)^2$$

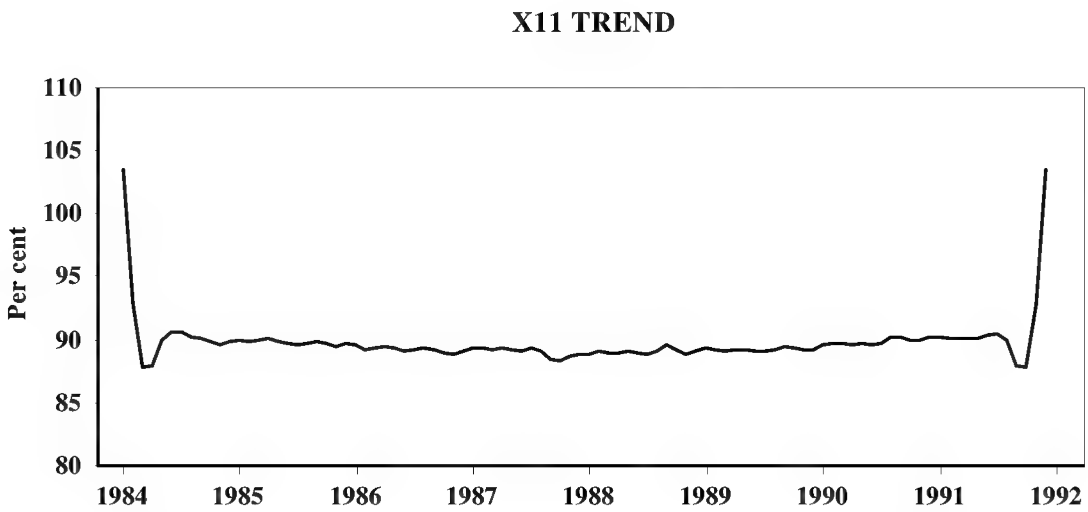
Hence the percentage of volatility of the levels of Monthly Retail Turnover due to the sampling process is approximated by

6.6 $$100 * \frac{\sigma_{es}^2(\tilde{e})}{\sigma_e^2(\tilde{e})}$$

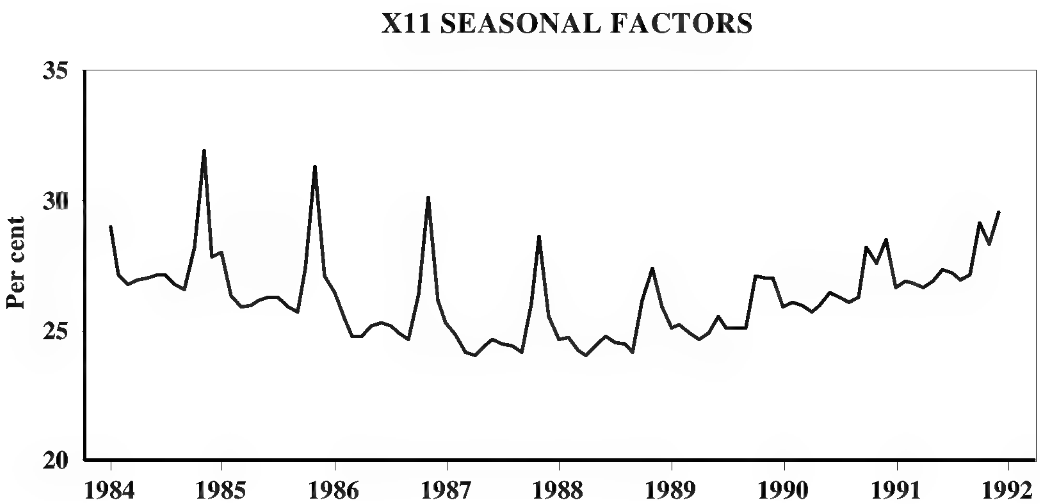
Graph 7 shows this percentage over time. In practice, the "real world" variance is almost certainly changing over time and there is plenty of empirical evidence that it is seasonal. It is a moot point whether the proportion should be to the actual observed residual or the residual modified for outliers. In the latter case the proportion would be much higher.

MONTHLY RETAIL TURNOVER
Standard error relative to original

GRAPH 1



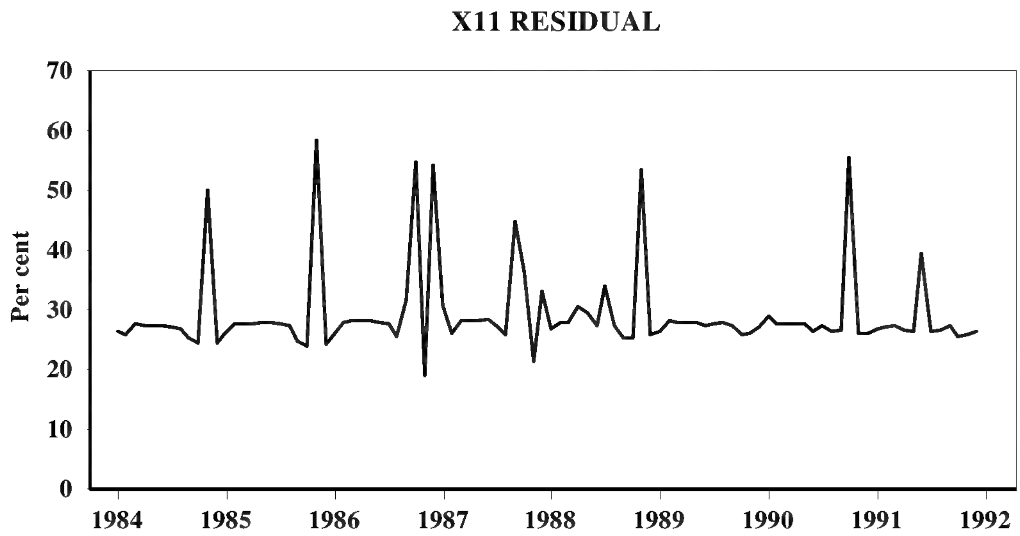
GRAPH 2



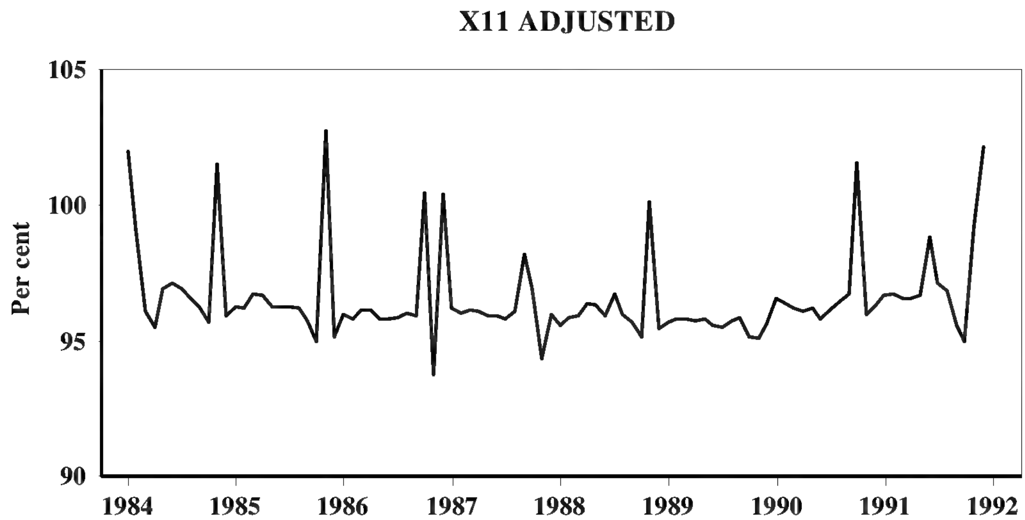
Note: 9 term Henderson, 3x5 seasonal, modification for extremes.

MONTHLY RETAIL TURNOVER
Standard error relative to original

GRAPH 3



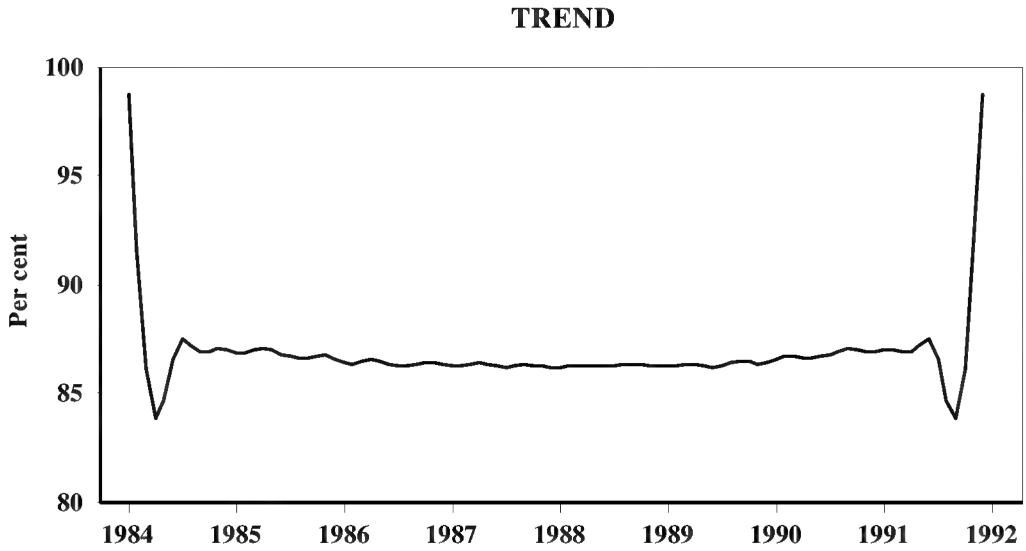
GRAPH 4



Note: 9 term Henderson, 3x5 seasonal, modification for extremes.

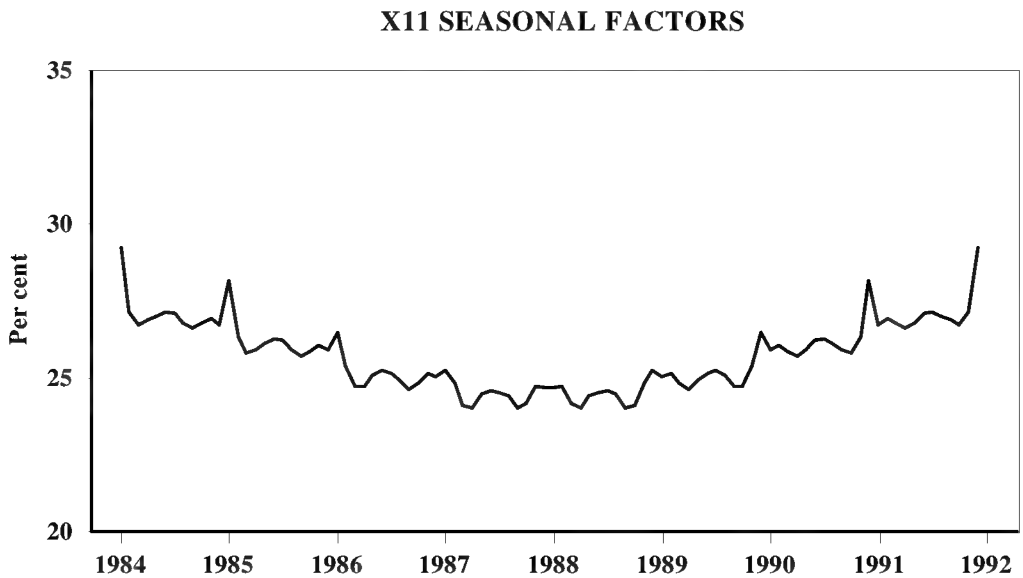
MONTHLY RETAIL TURNOVER
Standard error relative to original

GRAPH 5



Note : Published options, 13 term Henderson, no modification for extremes.

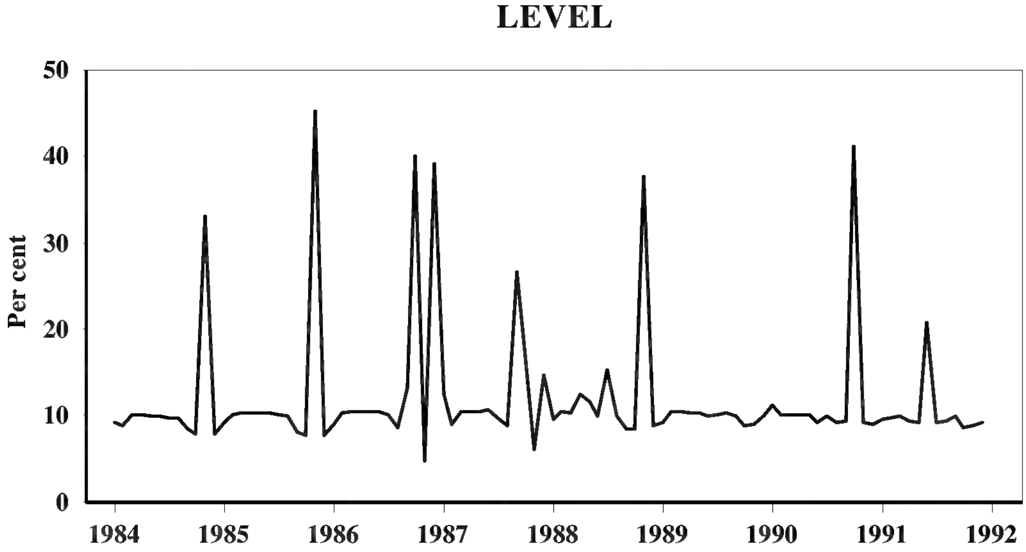
GRAPH 6



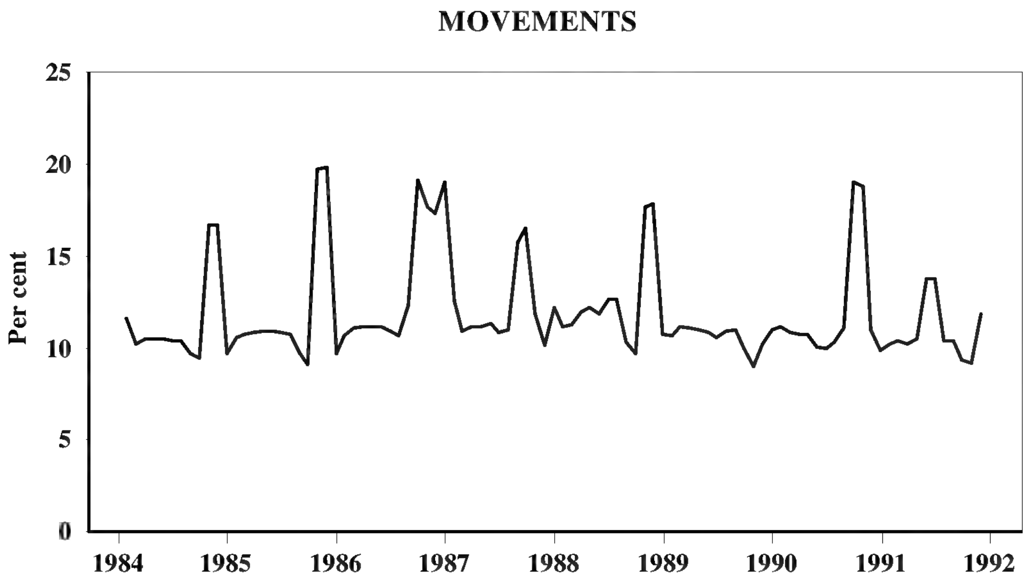
Note: 9 term Henderson, 3x5 seasonal, no modification for extremes.

MONTHLY RETAIL TURNOVER
Variance of sampling verses observed error

GRAPH 7



GRAPH 8



Note: 9 term Henderson, 3x5 seasonal, modification for extremes.
Observed variance assumed constant.

7. Variance of the movements.

These are estimated in a similar fashion to that of the levels shown in section 4. In this case define

7.1 $D^t = \hat{T} - B\hat{T}$ where B, the shift operator, shifts the rows of a matrix down one.

$$D^s = \hat{S} - B\hat{S}$$
$$D^e = \hat{E} - B\hat{E}$$

Then the approximate variances on the movements due to the sampling process is given by

7.2
$$\sigma^2_{es}((\tilde{t} - B\tilde{t})/B\tilde{t}) = diag(D^t * C * D^{t'})$$
$$\sigma^2_{es}((\tilde{s} - B\tilde{s})/B\tilde{s}) = diag(D^s * C * D^{s'})$$
$$\sigma^2_{es}((\tilde{e} - B\tilde{e})/B\tilde{e}) = diag(D^e * C * D^{e'})$$

Again, the variance of the change in the irregular component can be compared with the estimated variance of the observed change in the residual. This is given by

7.3
$$\sigma^2_e((\tilde{e} - B\tilde{e})/B\tilde{e}) = \frac{1}{n-1} \Sigma^n_{t=2}(\frac{e_t}{e_{t-1}} - 1)^2$$

Hence the percentage of volatility of the movements of Monthly Retail Turnover due to the sampling process is approximately given by

7.4
$$100 * \frac{\sigma^2_{es}((\tilde{e}-B\tilde{e})/B\tilde{e})}{\sigma^2_e((\tilde{e}-B\tilde{e})/B\tilde{e})}$$

Graph 8 shows this percentage over time.

8. Can better estimates of the "real" data be computed?

There are several approaches that can be used to provide better estimates of the "real" data. Unfortunately because there are so many sources of error any method proposed can always be criticised because of the assumptions used.

One approach is to use spectral analysis. For example if it could be ascertained, either with prior knowledge or empirical investigation, that the contribution of the variance due to sampling error was concentrated in a certain bandwidth and the "real world" and non-sampling error in a different bandwidth, then an appropriate filter could be designed using spectral analysis to remove/dampen the sampling error and leave the other error unchanged. More formally if the model for the observed data is

8.1
$$\tilde{y} = \tilde{t} + \tilde{s} + e\tilde{s} + e\tilde{r}$$

where

8.2 $\widetilde{es}$ is the sampling error and
$\widetilde{er}$ is the rest of the error.

Then if one could find a filtering matrix F such that

8.3 $$\widetilde{er} = F*(\widetilde{es}+\widetilde{er})+\widetilde{\varepsilon}$$

then it can be shown that a better estimate of the original data is given by

8.4 $$\widetilde{y}^{b} = \left(I-\left(\hat{E}+F*(I-\hat{E})\right)^{-1}*\hat{E}*(I-F)\right)*\widetilde{y}$$

8.5 $$= R*\widetilde{y}$$

There are several ways a suitable F matrix could be estimated. For example, if the sampling error is assumed to follow an autoregressive model then it can be written as

8.6 $$G*\widetilde{es} = \widetilde{\varepsilon}$$

Then if F depended on parameters they could be estimated by minimising

8.7 $$\widetilde{y}^{\prime}*(G*R)^{\prime}*(G*R)*\widetilde{y}$$

9. Sample design using time series characteristics.

If it is required to have an "optimal" sample allocation with respect to the decomposed data then clearly time series characteristics of the data must be taken into account. For example Australian Monthly Retail Turnover has allocation goals of minimising standard errors on movement and level for the original data. Generally if estimating the change in level between two time periods estimates of maximum precision are obtained by retaining the same sample on both occasions. For average or total values over a number of surveys non-overlapping units are selected. Clearly the trend level and movements are a weighted average of several survey values. If the goal was for a optimal sample for the trend component from the decomposition then the sample design/allocation may well be different.

In the example for Australian Monthly Retail Turnover the observed variance was assumed to be constant. In practice the variance may well be changing over time and is often observed to be seasonal.

It is possible to compute a time varying variance in a similar manner to that of the levels. For example, using a 5 year moving average and constant seasonality a model for the variance might be (where V is derived in a similar way to that of the levels in section 3)

9.1 $$\widetilde{\sigma}^{2}(\widetilde{e}) = V*\widetilde{e}^{2}$$

where

9.2 $$\widetilde{e} = \hat{E}*\widetilde{y}$$

In which case 9.1, could be used to assist in deriving an optimal sample allocation.

It should also be pointed out that autocorrelation in the residual due to sample design error (if strong enough) could be used to determine optimal filters for the time series decomposition.

10. Conclusions.

This paper demonstrates that a realistic approximation of X11, the widely used time series decomposition package, is possible. This provides a tool to allow extensive analysis of sample design and characteristics on time series decomposition. In particular, given a model for the covariance matrix of the sample design, and making certain assumptions the paper shows that
standard errors can be computed on components such as trend, seasonal and residual at all time points, and the proportion of variance due to sampling error can be estimated.

This paper only outlines the tools required for the analysis and further work needs to be completed on practical applications for individual surveys.

Currently, sampling design concentrates only on the original data. Given that the Australian Bureau of Statistics is giving more emphasis on the "trend" estimates it is not at all obvious that an "optimal" sample design for the original data will be "optimal" for the trend as estimated by the Australian Bureau of Statistics. This is clearly an area where further work is required to integrate the sample design and decomposition process.

REFERENCES

ABS Catalogue No 8501.0, *Retail Trade Australia*, April (1993).

Box G.E.P., Jenkins G.M., *Time series analysis forecasting and control*, 2nd ed, San Francisco, Holden Day, (1976).

Butter F.A.G., Mourik T.J., *Seasonal adjustment using structural time series models: An application and comparison with census X-11 method*, Journal of Business and Economic Statistics, Vol 8, No 4, October (1990), 385-393.

Cleveland W.P., Tiao G.C., *Decomposition of seasonal time series: A model for the census X-11 program, Journal of the American Statistical Association*, Vol 71, No 355, September (1976), 581-587.

Cleveland R.B., Cleveland W.S., McRae J., Terpenning I., *STL: A Seasonal-Trend decomposition procedure based on Loess*, Journal of Official Statistics, Vol 6, No 1, (1990), 3-73.

Dagum E.B., Chhab N., Chiu K., *Linear filters of the X11-ARIMA method*, Statistics Canada, Working Paper No BSMD-93-008E, (1993).

Hausman J.A., Watson M.W., *Errors in variables and seasonal adjustment procedures*, Journal of the American Statistical Association, Vol 80, No 391, (1985), 531-539.

Maravall S., Pierce D.A., *On structural time series models and the characterisation of components*, Journal of Business and Economic Statistics, 3, (1985), 350-355.

SAS Institute Inc., *SAS System Version 5*, Cary, NC, SAS Institute Inc, (1985).

Shiskin J., Young A.H., Musgrave J.C., *The X11 variant of the Census Method II: Seasonal adjustment program*, Technical Paper 15, US Department of Commerce, Bureau of the Census, (1967).

Steel D.G., de Mel R.G., *The contribution of sampling error to the variability of statistical series*, Unpublished, Australian Bureau of Statistics, (1987).

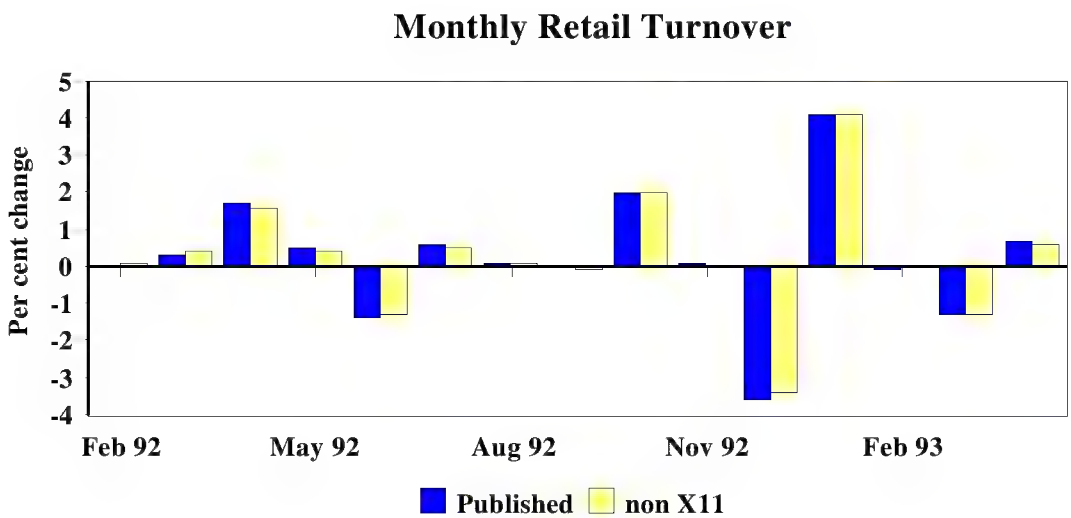
Young A.H., *Linear approximation to the Census and BLS seasonal adjustment methods*, JASA, (1968), 445-471.

Young A., Estimating trading-day variation in monthly economic time series, Technical Paper No 12, US Department of Commerce, Bureau of Census, (1965).

COMPARISON OF PUBLISHED SEASONAL ADJUSTMENT to APPROXIMATION OF X11

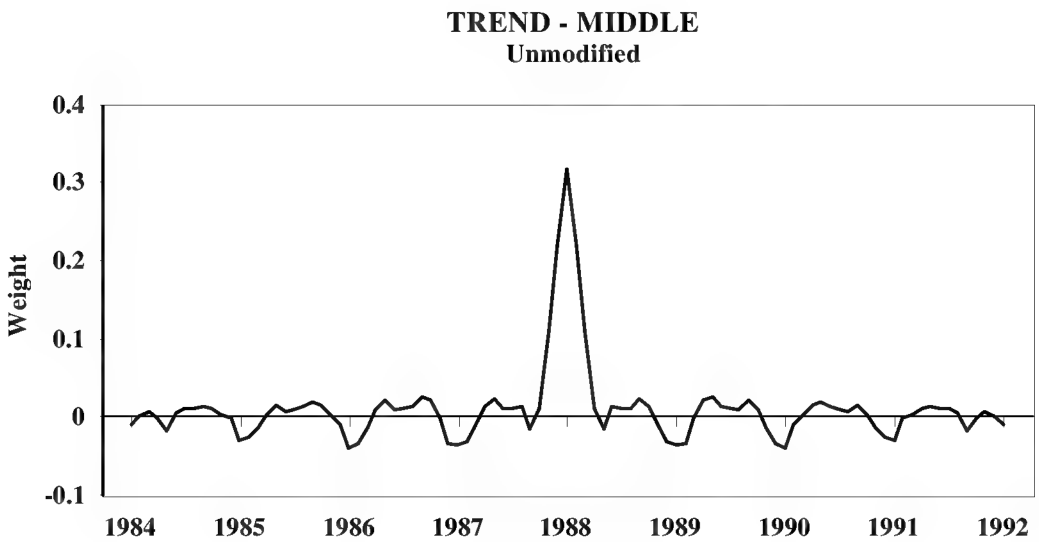
GRAPH A1

GRAPH A2

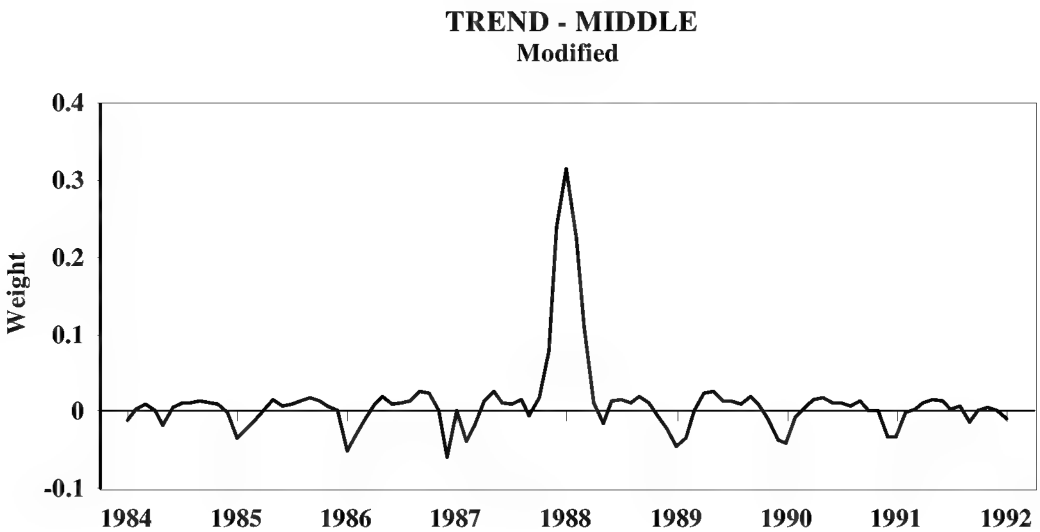


MONTHLY RETAIL TURNOVER
Filter approximation to X11

GRAPH A3



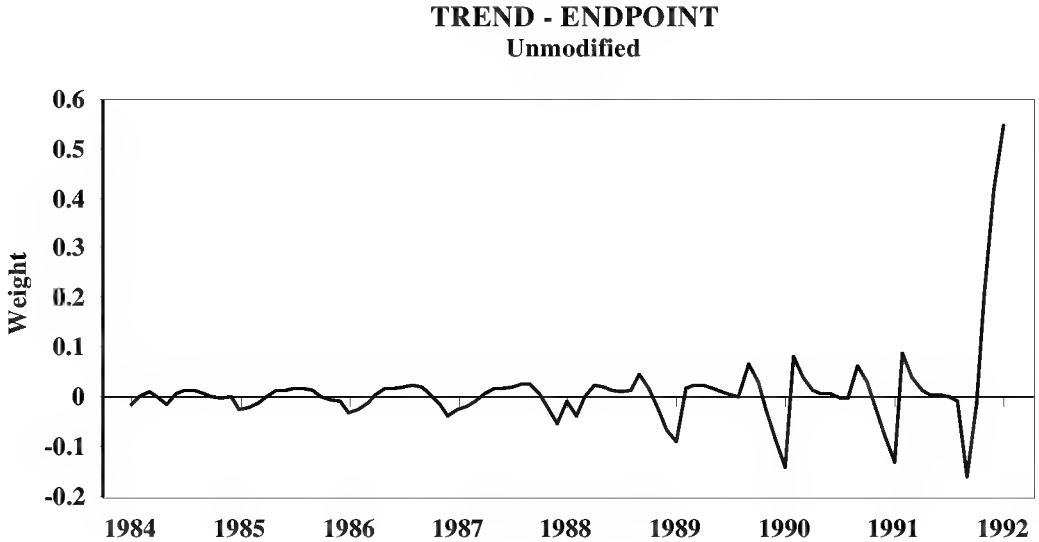
GRAPH A4



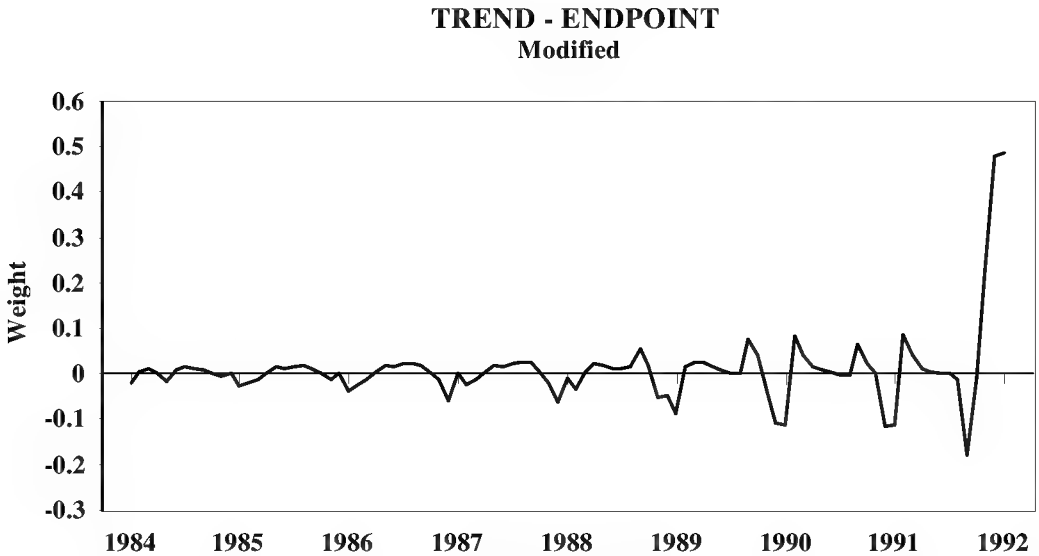
Note : 9 term Henderson, 3x5 seasonal.

MONTHLY RETAIL TURNOVER
Filter approximation to X11

GRAPH A5



GRAPH A6

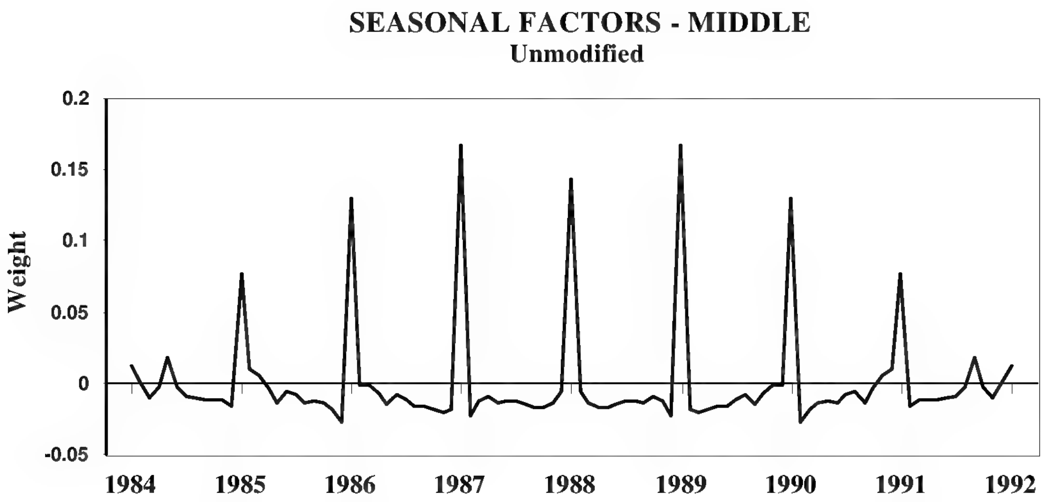


Note : 9 term Henderson, 3x5 seasonal

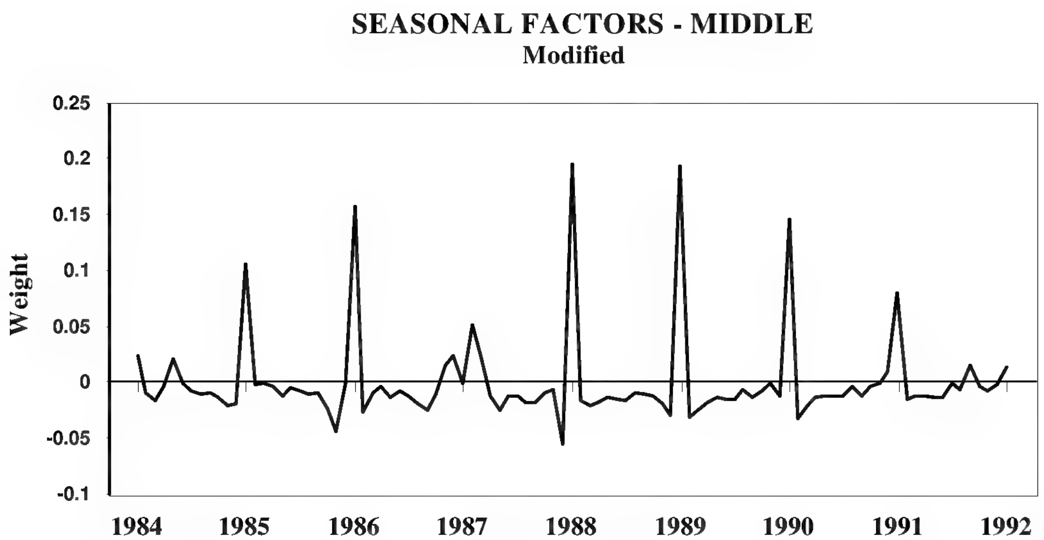
.

MONTHLY RETAIL TURNOVER
Filter approximation to X11

GRAPH A7



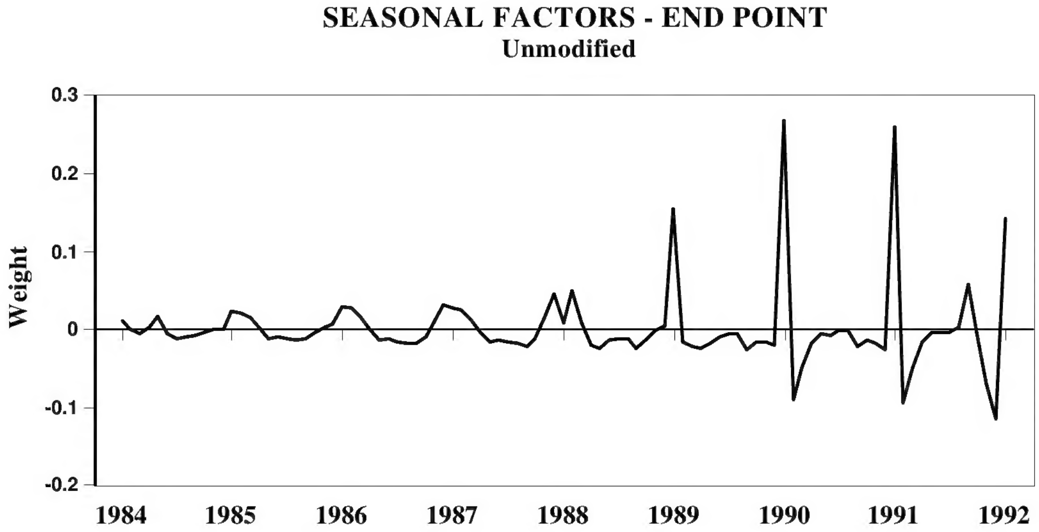
GRAPH A8



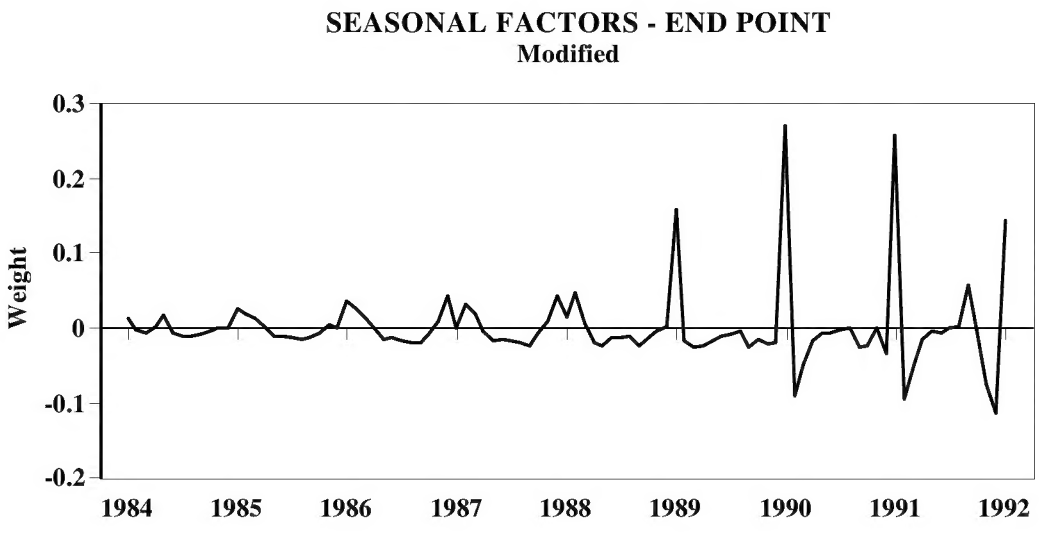
Note : 9 term Henderson, 3x5 seasonal.

MONTHLY RETAIL TURNOVER
Filter approximation to X11

GRAPH A9



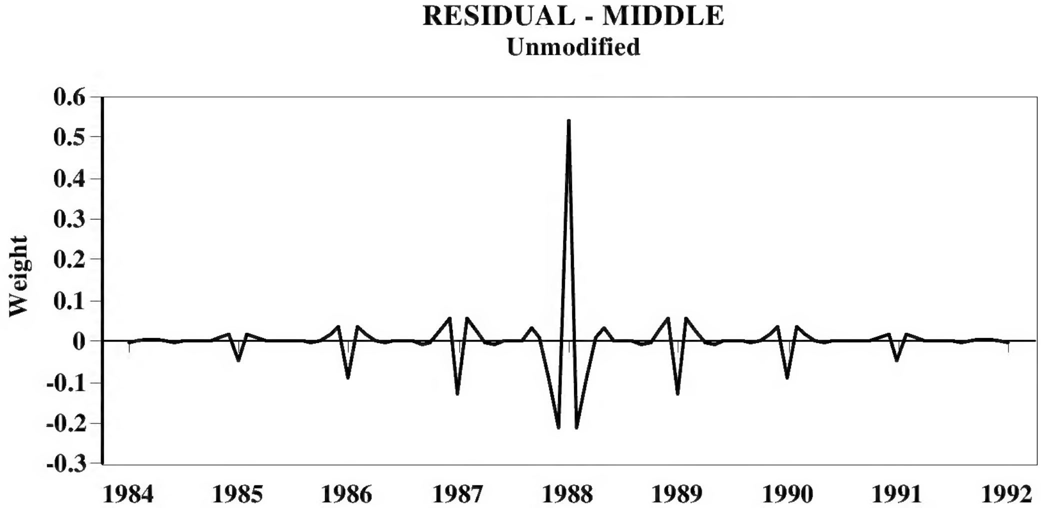
GRAPH A10



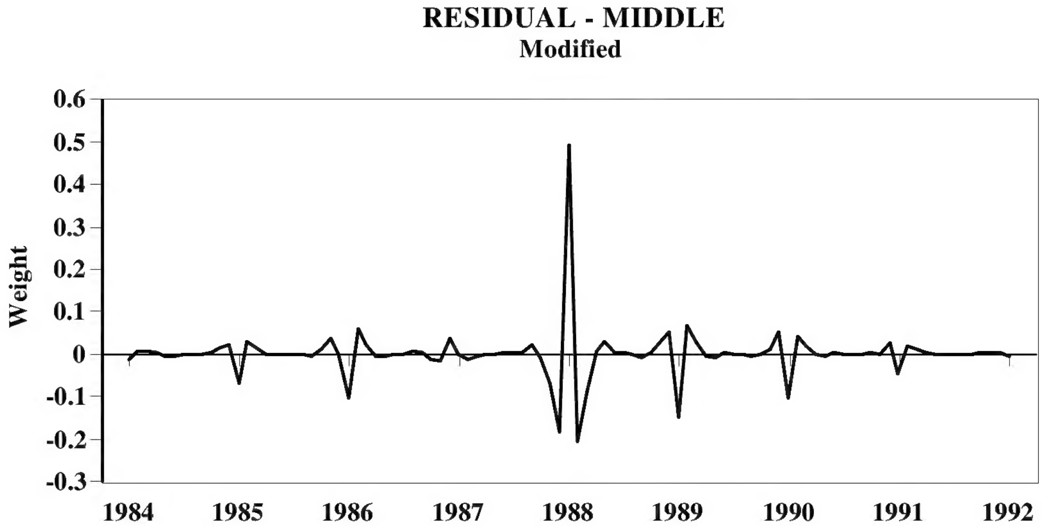
Note : 9 term Henderson, 3x5 seasonal.

MONTHLY RETAIL TURNOVER
Filter approximation to X11

GRAPH A11



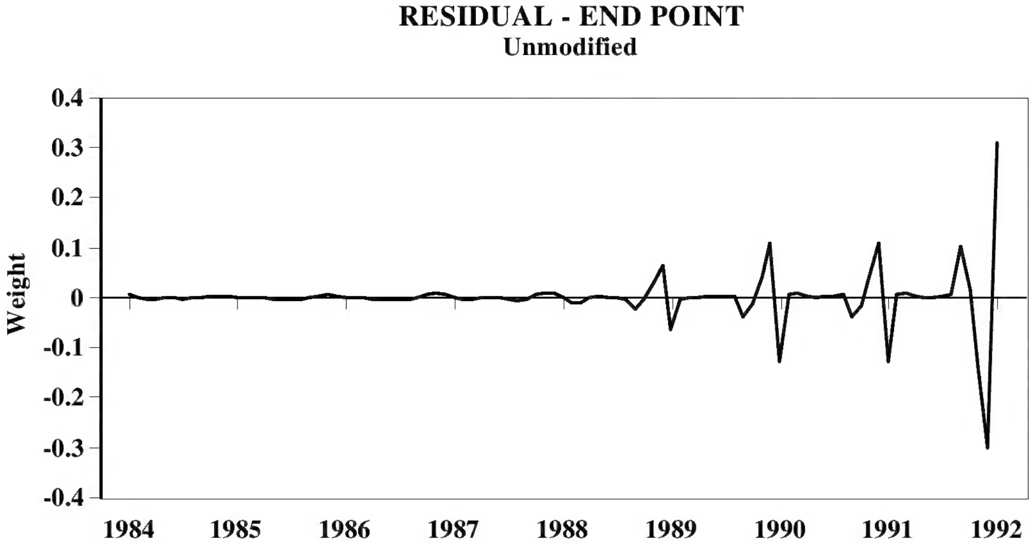
GRAPH A12



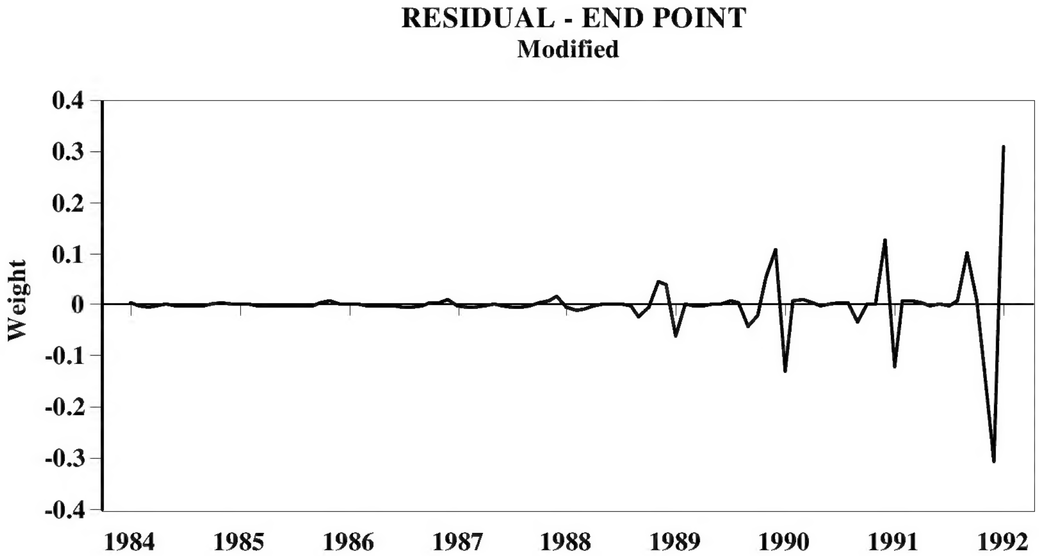
Note : 9 term Henderson, 3x5 seasonal.

MONTHLY RETAIL TURNOVER
Filter approximation to X11

GRAPH A13



GRAPH A14



Note : 9 term Henderson, 3x5 seasonal.

Contents

INQUIRIES

For further information about the contents of this Working Paper, please contact the author:

Andrew Sutcliffe - Canberra (06) 252 7646 *(telephone) or* (06) 253 1093 *(facsimile).*

For information about the Working Papers in Econometrics and Applied Statistics, *please contact the Managing Editor,* Geneviève Knight *on* Canberra(06) 252 6427 *(telephone) or* (06) 253 1093 *(facsimile), or write care of* Econometric Analysis